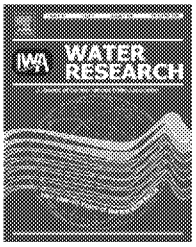
WATER RESEARCH 41 (2007) 4299–4308

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/watres



Probability of detecting and quantifying faecal contaminations of drinking water by periodically sampling for *E. coli*: A simulation model study

J. Hein M. van Lieverloo[a,*,1], *George A.M. Mesman*[a], *Geo L. Bakker*[b], *Paul K. Baggelaar*[a], *Anas Hamed*[b], *Gertjan Medema*[a]

[a]*Kiwa Water Research, Groningenhaven 7, 3433 PE Nieuwegein, The Netherlands*
[b]*Vitens, Boogschuttersstraat 29a, 7324 AE Apeldoorn, The Netherlands*

ARTICLE INFO

Article history:
Received 15 November 2006
Received in revised form
29 May 2007
Accepted 3 June 2007
Available online 12 June 2007

Keywords:
E. coli
Faecal contamination
Drinking water
Detection
Sampling
Simulation

ABSTRACT

Drinking water supply companies monitor the presence of *Escherichia coli* in drinking water to verify the effectiveness of measures that prevent faecal contamination of drinking water. Data are lacking, however, on the sensitivity of the monitoring programmes, as designed under the EU Drinking Water Directive. In this study, the sensitivity of such a monitoring programme was evaluated by hydraulic model simulations of contamination events and calculations of the detection probability of the actual sampling programme of 2002. In the hydraulic model simulations of 16-h periods of $1\,l\,h^{-1}$ ingress of untreated domestic sewage, the spread of the contamination through the network and the *E. coli* concentration dynamics were calculated. The results show that when large parts of the sewage reach reservoirs, e.g. when they originate from the treatment plant or a trunk main, mean detection probabilities are 55–65%. When the contamination does not reach any of the reservoirs, however, the detection probability varies from 0% (when no sampling site is reached) to 13% (when multiple sites are reached). Mean detection probabilities of nine simulated ingress incidents in mains are 5.5% with an SD of 6.5%. In reality, these detection probabilities are probably lower as the study assumed no inactivation or clustering of *E. coli*, 100% recovery efficiency of the *E. coli* detection methods and immediate mixing of contaminations in mains and reservoirs.

The described method provides a starting point for automated evaluations and optimisations of sampling programmes.

1. Introduction

One of the major responsibilities of drinking water supply companies is to prevent faecal contamination of drinking water. Barriers in treatment remove faecal contaminations from surface water or contaminated ground water. Barriers in infrastructure prevent contamination from the vicinity (ingress via e.g. leaks) or connected buildings (backflow, backpressure). The maintenance of overpressure in the distribution system is a major assisting barrier for entrance of contaminations. Furthermore, guidelines for hygiene prevent faecal contaminations during operations (Van Lieverloo et al., 2006).

In many countries, water companies strive to maintain a disinfectant residual in the distributed water to inactivate

*Corresponding author. NIZO, P.O. Box 20, 6710 BA Ede, The Netherlands. Tel.: +31 318 659 558; fax: +31 318 650 400.
E-mail address: Hein.van.Lieverloo@nizo.nl (J.H.M. van Lieverloo).
[1] Present address: NIZO Food Research, Kernhemseweg 2, 6718 ZB Ede, The Netherlands.
0043-1354/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.watres.2007.06.003

pathogens that inadvertently may have entered the distribution system. In The Netherlands and in large parts of Germany, this disinfectant residual is absent or low and the water companies rely on other barriers to be effective in preventing contaminations. The effectiveness of disinfectant residuals to inactivate pathogens is considered too limited to rely on and the negative effects of disinfectants on health (by-products) and taste are considered too high (Van der Kooij et al., 1999; Hambsch, 1999; Payment, 1999).

Testing for the presence of *Escherichia coli* in drinking water is the most common verification of the effectiveness of contamination prevention. It is retrospective, as the samples are collected in the houses of consumers and provide information about the presence or absence of a faecal contamination only after water is consumed. The presence of *E. coli* is indicative of a faecal contamination and the highly probable presence of pathogens. The absence of *E. coli*, however, is no guarantee for the absence of these pathogens, as the retention and survival of *E. coli* and pathogens in many environments, including drinking-water distribution systems, is not proportional (Ashbolt et al., 2001). In a significant proportion (31%) of outbreaks of waterborne illness in (chlorinated) public water supply systems in the US that occurred between 1991 and 1998, no coliforms have been detected in the routine monitoring programme (and the additional samples taken in the cause of the outbreak investigation) (Craun et al., 2003). In chlorinated systems, this can be the result of more rapid inactivation of coliforms than pathogens, as suggested by the low rate of *E. coli* reported during *Cryptosporidium* and *Giardia* outbreaks (26%) as opposed to outbreaks of bacteria, viruses and unknown agents (88%; Craun et al., 2003). A major drawback of *E. coli* monitoring, as well as monitoring for other variables, is the incomplete coverage of the sampling programme. The characteristics determining the detection probability are

- spatial density of the sampling sites,
- frequency of sampling and
- detection limit per sample (caused by analysed volume and *E. coli* viability).

No detection programme is able to detect all local and/or temporary faecal contaminations, unless continuously operating sensors are installed on every single main or house. No information is available about the sensitivity of the current distributed water quality monitoring programmes. The objective of this study was to evaluate the sensitivity of a current *E. coli* monitoring programme, designed in accordance with the EU Drinking Water Directive (European Council, 1998), to detect a faecal contamination.

To achieve this objective, a hydraulic contamination model simulation was developed to calculate the spread of the *E. coli* concentration in a distribution system after contamination with sewage. The actual monitoring programme was superimposed on the calculated concentrations to assess the probability of detecting these faecal contaminations.

2. Materials and methods

2.1. *Hydraulic modelling and contamination spreading*

Many water companies in The Netherlands use ALEID® for hydraulic modelling of their distribution systems. ALEID is an EPANET-based (Rossman et al., 1994; Vasconcelos et al., 1997; EPA, 2000) network analysis programme. The model uses Darcy–Weisbach/Colebrook for the calculation of pressure drops and volume flows. Within ALEID, the spreading of contaminations (faecal and other) can be modelled. Contaminations can be considered as conservative or as changing with a first-order reaction.

2.2. *Case study distribution system and hydraulic model*

The distribution system of the town and nearby rural areas of Almelo, The Netherlands, was selected because a detailed hydraulic model was readily available. The model (nodes and pipes) was derived from the GIS of the water company and loaded with very accurate spatial demand data from a fully metered situation. The diurnal demand pattern for all demands is derived from a complete water balance of the distribution system.

The distribution system is supplied by treatment plant Wierden, collecting anoxic water from 27 wells and treating this by two stages of rapid sand filtration in series (interrupted by aeration). A water tower (600 m^3) is situated in the old centre of the town and a bottom reservoir (2500 m^3) is situated in the east side. The water tower has a semi-free water level; the inlet is closed when the water tends to overflow at times of low demand. The bottom reservoir has a daily cycle of filling (during the night) and suppletion (25% during the day) to the system.

Fig. 1 shows the outline of the hydraulic model. Key characteristics of the distribution system and the hydraulic model are presented in Table 1.

2.3. *Validation/calibration of the model*

The validation of the model is executed automatically by software that checks anomalies such as not existing diameters, not existing diameter–*k*-value combinations, strange transitions in pipes and non-connections where there are connections expected.

A hydraulic calibration is executed for a dataset of 10 November 2002. The model is calibrated by comparing the supplied and calculated volume flows of the production plant, the reservoir and the water tower. For this set of measurements, 79% of the calculated values lies within 5% of the total flow (water balance of the area) and 96% lies within 10%. For the flow from the production plant, 67% lies within 10% of the measured flow and 88% lies within 20%. The pressures in the network were measured and compared to the calculated pressures. The match of the compared pressures was 83% within 5% of the measured values and 96% within 10%.

There was no need for calibration of the water quality part of the model, e.g. by trace studies, as the contamination of the network was considered conservative.

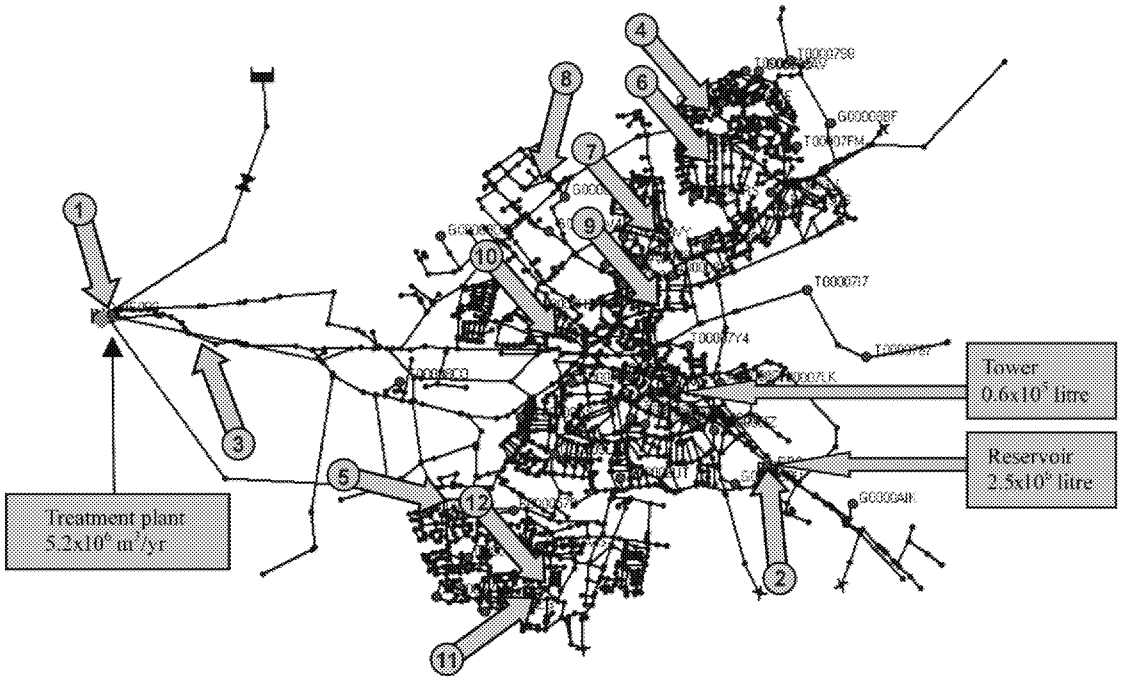


Fig. 1 – Outline of the hydraulic model of the distribution system of the town of Almelo supplied by treatment plant Wierden. The 44 sampling sites in the distribution system are text-labelled and indicated with white dots (red dots in the online version). The 12 simulated contamination sites are indicated with numbered arrows.

Table 1 – Key characteristics of the case study distribution system and the hydraulic model

Variable	Characteristic
Yearly demand of Almelo (2003) (72,000 inhabitants, 20,300 connections)	$4.23 \times 10^6\,m^3$
Domestic	$2.86 \times 10^6\,m^3$
Small industrial	$0.64 \times 10^6\,m^3$
Industrial	$0.74 \times 10^6\,m^3$
Supply to other distribution systems	$1.00 \times 10^6\,m^3$
Total supplied by treatment plant Wierden	$5.23 \times 10^6\,m^3$
Total length of the network	332 km
Internal diameter <160 mm	244 km
Internal diameter >160 mm	98 km
Nodes in the model	3310
Pipes in the model	3866
Average pressure	320 kPa (32 mwc)

2.4. *Simulated faecal contamination incidents*

For each of the simulated incidents, the same scenario for faecal contamination was selected; for 16 h untreated domestic sewage flows in, with a flow of $1\,l\,h^{-1}$. The 16-h contamination scenario was chosen as a 'worst case' scenario, simulating a relatively large, though conceivable contamination. A concentration of *E. coli* of $1 \times 10^8\,CFU\,l^{-1}$ (total of $1.6 \times 10^9\,CFU$) was used in the calculations, based on data collected at two large sewage treatment plants (Medema et al., 2001). On every contamination site, a pump and a supply main were entered into the model to model the flow of sewage. *E. coli* was considered to be conservative (no inactivation). The resulting *E. coli* concentration dynamics on all 47 sampling sites were calculated for hourly intervals in a period of 1200 h (50 days) after each contamination, starting at midnight. Immediate complete mixing of sewage with the water in the mains, as well as in the reservoirs, was assumed, as the model does not include computational fluid dynamics in large mains and reservoirs. Also, complete suspension of *E. coli* was assumed.

2.5. *Selection of sites for simulated contaminations*

Simulated contaminations were created on 12 sites; i.e. the finished water of the treatment plant, one of the trunk mains, one of the finished water reservoirs and nine distribution mains in different areas of the distribution system. The distribution mains were selected throughout the city, on sites varying from mains supplying larger areas to peripheral mains. In this stage of the investigations, it was not possible to evaluate detection probabilities for a large number of randomly selected contamination sites. The sites are indicated in the outline of the hydraulic model in Fig. 1.

2.6. *Case study sampling programme for E. coli in drinking water*

The sampling sites and dates actually applied in 2002 in the case study distribution system were used during the evaluation. The sites are indicated in the map in Fig. 1. The sampling programme is fully complying with the EU Drinking water directive (European Council, 1998). According to this directive, a minimum of 39 check monitoring samples and 15 audit

Table 2 – Key statistics of the 2002 sampling program in the case study distribution system

	Sites	Samples	Sampling interval (days)			
	N	N	Mean	SD	Median	Max
Treatment plant	1	53	6.9	1.8	7	12
Reservoirs	2	22	34.5	15.5	33	63
Buildings	44	176	91.3	23.7	88	150

samples should be collected, distributed in time and location as equally as possible. The number of samples collected in buildings is over three times the required number and the total number of samples, including samples collected from the treatment plant and distribution reservoirs, is almost five times the required number (Table 2). The samples from buildings usually are collected weekly, not daily, in clusters of three to four, together with weekly samples from treatment plants.

2.7. *Calculating detection probabilities from concentrations dynamics*

For each of the 12 simulated faecal contamination incidents, the resulting *E. coli* concentrations dynamics of all 47 sampling sites were evaluated. The probability of detecting *E. coli* on a single sampling site is the period during which *E. coli* concentration is $\geqslant 10\,\mathrm{CFU\,l^{-1}}$ divided by the period between sampling moments and $10\,\mathrm{CFU\,l^{-1}}$ is the detection limit of the analysis (as 100 ml of each sample is tested). The probability of detecting a faecal contamination, however, is the combined detection probabilities of all sampling sites. This combination is the sum (of all hours) of the maximum (of all sites) of detection probabilities per hour after the contamination. The height of these combined probabilities largely depends on the date of contamination relative to the sampling dates. Therefore, for every day of 2002, the calculated concentrations of *E. coli* on each sampling site on the first sampling day following the contamination were evaluated for the relative number of hours of that day during which this concentration was $\geqslant 10\,\mathrm{CFU\,l^{-1}}$. This was evaluated assuming samples are taken any time of the day (i.e. 24 hourly moments), only at night-time (i.e. 8 hourly moments from 00.00 to 07.00 h), only at daytime (i.e. from 08.00 to 15.00 h) and only at evening time (i.e. from 16.00 to 23.00 h).

2.8. *Simulating different contamination levels*

In the actual model calculations, a contamination with 16 l of sewage was simulated (1.6×10^9 CFU of *E. coli*). It was possible to simulate the detection probabilities of larger and smaller contaminations without repeated (time-consuming) hydraulic modelling of concentrations dynamics. By changing the detection limit of the analysis during calculations of detection probabilities, contaminations with 160, 1.6 and 0.16 l of sewage were calculated (using detection limits of 1, 100 and $1000\,\mathrm{CFU\,l^{-1}}$, respectively, instead of $10\,\mathrm{CFU\,l^{-1}}$).

3. Results

3.1. *Spreading*

In most of the 12 simulated contaminations, the resulting geographical spreads were in line with expectations, viz. contaminations occurring in the periphery tend to reach smaller parts of the distribution system (Fig. 2). In two of the nine simulated contamination incidents in mains (simulations 6 and 8), none of the 44 sampling sites and two distribution reservoirs were contaminated, rendering the detection probabilities of contaminations on these sites zero. In one simulated incident (simulation 11), six of the 44 sampling sites were contaminated in concentrations that were only detectable when 1600 l of sewage or more had entered the distribution main. In some mains contaminations (simulations 5, 10 and 12) an unexpected flow to either or both reservoirs was found, increasing probabilities of detecting severe contaminations. From the reservoir or reservoirs, larger areas of the distribution system were contaminated, although concentrations on most sampling sites were below the detection limit ($10\,\mathrm{CFU\,l^{-1}}$) within 1 week.

3.2. *Dilution and flush-out*

In the model, mixing of the contamination with the drinking water at the contamination site occurred completely and immediately upon entrance. The *E. coli* concentrations and detection probabilities decreased with increasing flow in the contaminated main or increasing volume of the reservoir. Furthermore, the contamination peak widens and lowers as water flows toward the periphery as it flows to a site via more than one trunk main and a web of mains, allowing fresh, uncontaminated water to mix with contaminated water.

Flush-out normally only occurs via the taps in the connected buildings. Fig. 3A shows the *E. coli* concentrations dynamics in these reservoirs after a faecal contamination of the treatment plant. *E. coli* was detectable for several days in these reservoirs. The flush-out of a contamination in any of the reservoirs was only minimal, a result of their function as auxiliary supplies. In the same simulated contamination event, *E. coli* could be detected for a couple of days on a sampling site (partly) supplied by either of the auxiliary reservoirs (Fig. 3B). These figures also indicate that a more sensitive monitoring programme (larger volumes) would have a higher probability of detecting this event.

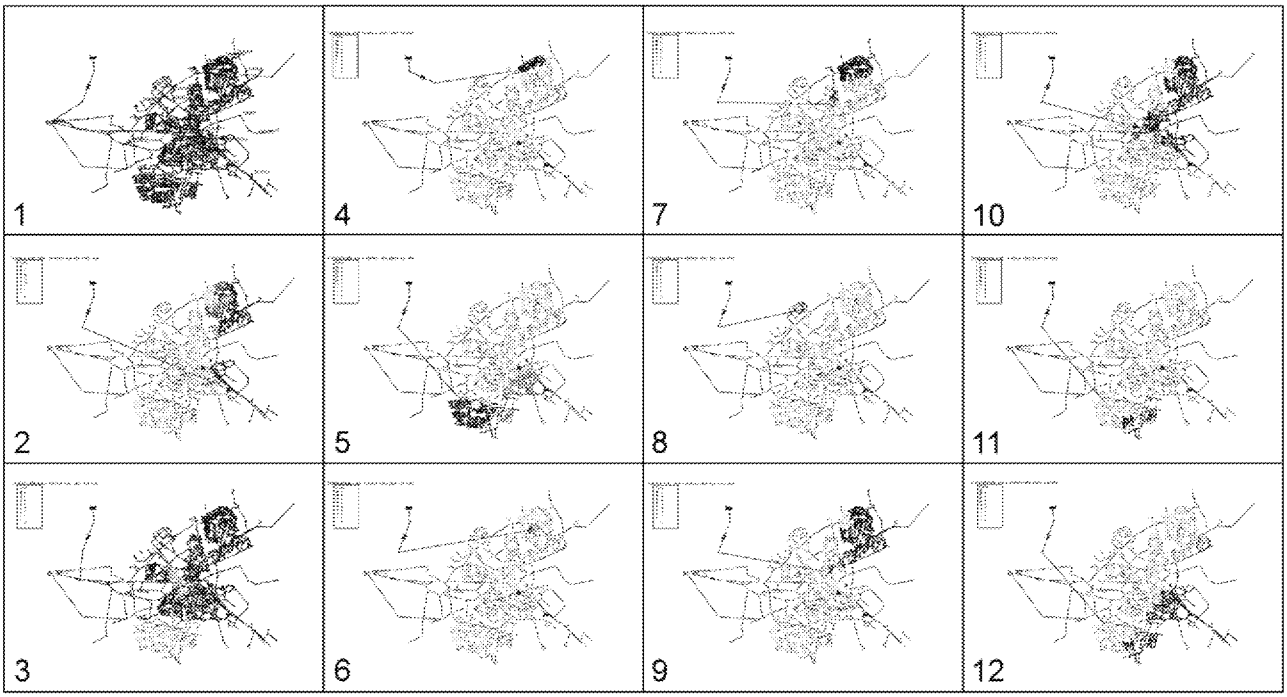


Fig. 2 – Contaminated areas (black, concentrations >0.01 CFU l^{-1}) of the model system per simulated contamination incident, with 16 l of sewage (1.6×10^9 CFU *E. coli*) in 16 h, assuming no inactivation of *E. coli*. Incident numbers correspond with numbers in Fig. 1.

The detection probabilities increased with the volume of sewage that entered, but this increase was limited by the flush-out period. The flush-out period was only 16 h in the finished water of the treatment plant (Fig. 3A), and was up to (only) several days in mains not supplied by any of the reservoirs (Fig. 3B), independent of the volume of sewage entering the system. At the highest contamination level, the calculated mean detection probability was approx. 11% in the finished water (Fig. 4A), i.e. slightly over the probability of taking a sample during the 16 h of the contamination event when sampling with a mean interval of 6.9 days (165 h, Table 2). The maximum mean detection probability is approx. 4% on a sampling site not supplied by any of the reservoirs (Fig. 4B), i.e. approx. 4 days between samples taken with mean intervals of 91.3 days.

3.3. *Effect of spreading sampling dates*

The combined probability of detecting a faecal contamination (16 l of sewage) of the finished water leaving the treatment plant on either of the sampling sites on distribution mains was higher (approx. 50%, Fig. 5) than in the finished water itself (approx. 11%) or on any of the individual sampling sites (Fig. 4). This was caused by the spreading of the sampling dates of the 44 sampling sites in the distribution network monitoring programme.

3.4. *Effect of the location of the contamination site*

When all or a large portion of the entered sewage reached either or both of the reservoirs, the initial *E. coli* concentrations in these reservoirs were high. Therefore, the combined detection probabilities of contaminations at the treatment plant, the selected trunk main and the bottom reservoir were high (means between 55% and 65% for contaminations with 16 l of sewage, Fig. 6).

When only a small part of the entered sewage reached either or both of the reservoirs, *E. coli* concentrations more quickly dropped to levels below the detection limit, hence lowering detection probabilities. In simulations 5, 10 and 12, a small portion of the contamination flowed from the contaminated distribution mains into either or both of the reservoirs (combined probability approx. 6%, 14% and 3%, respectively, for contaminations with 16 l of sewage, Fig. 6).

When a contamination of a distribution main did not reach either of the reservoirs, but did reach sampling sites connected to mains, the detection probability was even lower or zero, depending on the severity of the contamination (simulations 4, 7, 9 and 11, combined probability approx. 6%, 6%, 13% and 0%, respectively, for contaminations with 16 l of sewage, Fig. 6). When neither of the detection sites was reached by the contamination, the detection probabilities of any contamination were zero (simulations 6 and 8, Fig. 6).

The means of the simulated contaminations of 16 l in 16 h into the nine selected distribution mains varied between 0% and 14% (SD 28%), resulting in an overall mean and SD of these means of 5.4%±5.3% (median 6%).

3.5. *Effect of sampling moment*

In general, the detection probabilities of contaminations in mains were slightly higher during the evening (16.00–00.00 h),

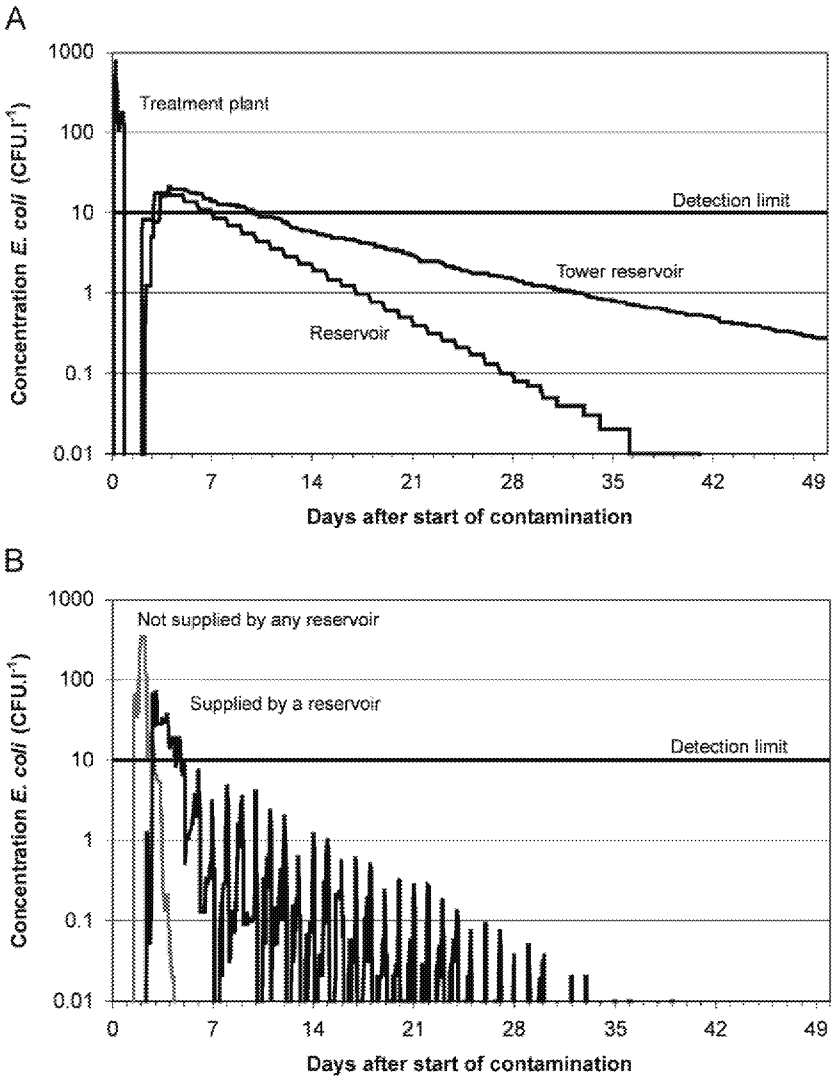


Fig. 3 – Concentration dynamics of *E. coli* after a simulated contamination of the finished water of the treatment plant of the model system with 16 l of sewage (1.6×10^9 CFU *E. coli*) during 16 h, assuming no inactivation of *E. coli*. Plot A shows the concentration dynamics in the finished water of the treatment plant and in the reservoirs (assuming immediate complete mixing after entry of the faecal contamination). Plot B shows the concentrations on two sampling sites in buildings connected to mains, where one is not supplied by any of the reservoirs, and the other is (intermittently and partly) supplied by a reservoir. The detection limit is $10\,l^{-1}$.

followed by the daytime (08.00–16.00 h) and the night (00.00–08.00 h). The mean absolute differences between sampling periods were $1.4 \pm 0.8\%$ for 16 l contaminations of the treatment plant, bottom reservoir or trunk main (relative difference range 0.2–5%, median 2.5%). The mean absolute differences were $0.6 \pm 0.5\%$ for the six detectable 16 l contaminations of mains (relative difference range 0.8–80%, median 5%). None of the differences were statistically significant (paired *T*-tests).

3.6. *Optimisation of the monitoring programme*

The effect of more homogeneous spreading of sampling dates over the year was evaluated for the same contamination incidents. In the actual sampling programme, the mean interval between samples in any of the sampling sites of connected buildings was 2.1 days with an SD of 5.0 days (due to clustering of samples on a single day per week). In a fully homogeneous programme, the 176 samples are collected on intervals of 2 (in some cases 3) days, including Saturdays and Sundays, reducing the SD to 0.25 days. Samples of finished water are collected every 7 days, reducing the SD from 1.8 days to 0 days. Samples in the two reservoirs are taken every 15 or 16 days by spreading the monthly samples, reducing the mean and SD from 17 ± 12 days (22 samples) to 15 ± 0.7 days (24 samples). Due to this homogenisation, the mean detection probability of contamination in mains increases from 5.4% to approx. 7%. The detection probabilities of contaminations in the finished water, the trunk main or the bottom reservoirs increases from 55–65% to 74–82%.

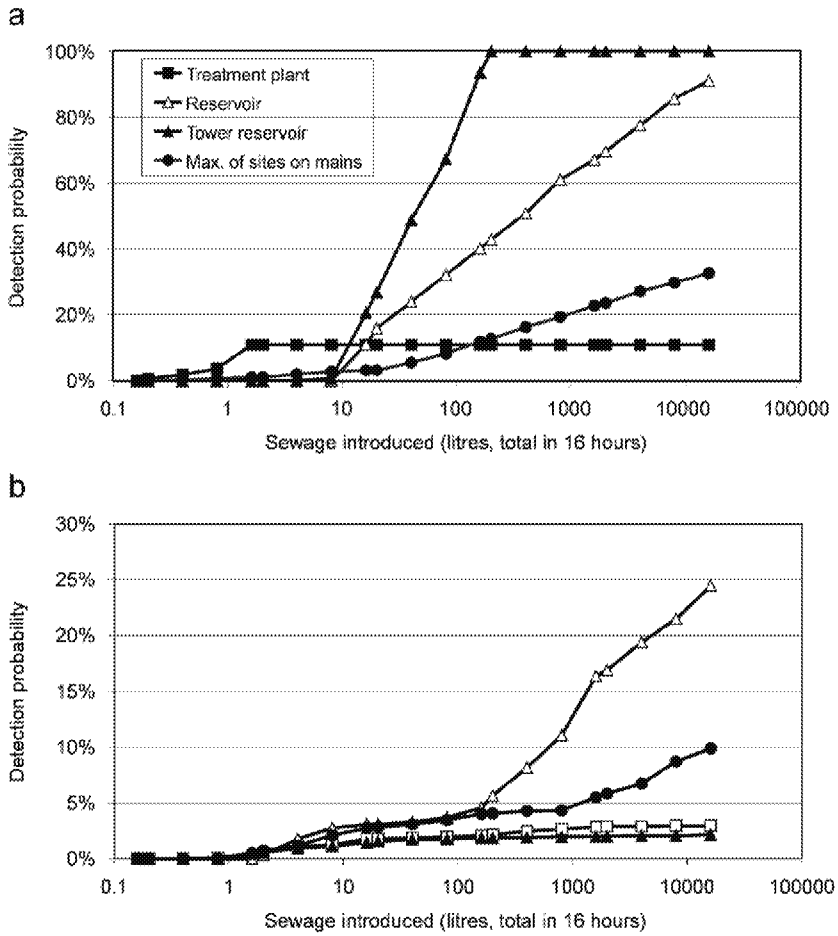


Fig. 4 – Mean detection probabilities on individual sampling sites after simulated contamination of the finished water leaving the treatment plant with different volumes of sewage (containing 1.0×10^8 CFU l^{-1} of *E. coli*) during 16 h after midnight, assuming no inactivation of *E. coli*. Plot A shows the detection probabilities in the treatment plant, the reservoirs and the maximum of the detection probabilities in buildings connected to mains. Plot B shows the detection probabilities in four individual buildings connected to mains. Note the difference in the scale of the y-axes.

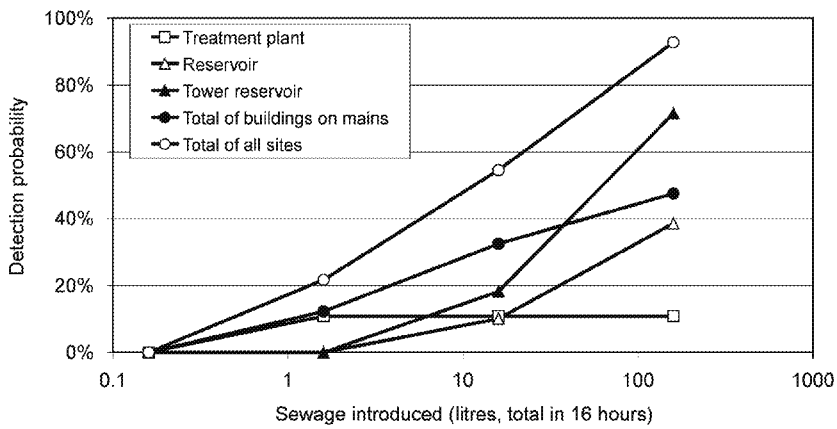


Fig. 5 – Mean detection probabilities after simulated contaminations, of increasing severity, of the treatment plant. Probabilities were plotted for finished water (labelled as 'Treatment plant') and the reservoirs ('Reservoir' and 'Tower reservoir'). Furthermore, the mean combined detection probabilities on all 44 sampling sites in buildings ('Total of buildings on mains') and all 47 sampling sites including finished water, reservoirs and buildings ('Total of all sites') were plotted. Simulations were executed with different volumes of sewage (containing 1.0×10^8 CFU l^{-1} of *E. coli*) during 16 h after midnight, assuming no inactivation of *E. coli*.

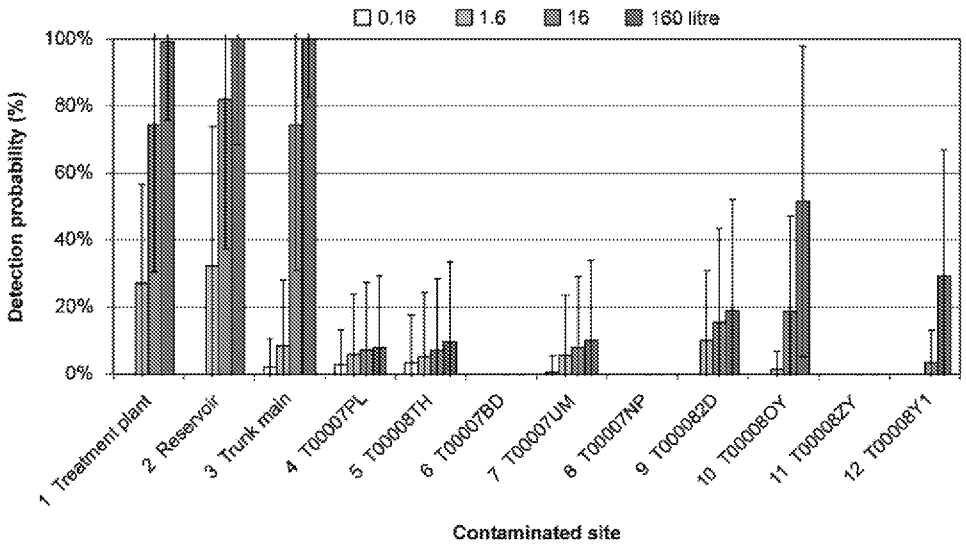


Fig. 6 – Means and standard deviations of the combined detection probabilities of sampling on all 47 sites (finished water, the two reservoirs and 44 buildings along the distribution network) of simulated contaminations on 12 different sites (each contamination at one site only) in the model distribution system with 0.16, 1.6, 16 and 160 l of sewage (each litre containing 1.0×10^8 CFU *E. coli*) during 16 h after midnight, assuming no inactivation of *E. coli*.

4. Discussion

The calculated probabilities of detecting a faecal contamination event in the distribution mains is low. A number of modelling choices and assumptions, as described in the methods, may have caused an overestimation of the probability of detecting the faecal contamination event:

- *E. coli* was assumed conservative (no inactivation). In reality, *E. coli* inactivation occurs, even without disinfectant residuals.
- The detection method was assumed 100% sensitive (no false negatives) while *E. coli* bacteria that are injured by stressors in the water environment may not be recovered by the selective culture methods used, even when resuscitation is used (Bjergbæk and Roslev, 2005).
- Complete mixing of *E. coli* bacteria from sewage in water was assumed, resulting in a random (Poisson) distribution of *E. coli* in water. In faecally contaminated (surface) water, the spatial distribution of faecal micro-organisms is usually more skewed due to clustering (Medema and Schijven, 2001).
- Immediate mixing in reservoirs, as assumed in the model, will not occur as concentrations will remain higher near the trunk mains both while filling as well as draining the reservoirs. Therefore, flush-out of the contamination will occur faster. Computational fluid dynamic modelling will result in more realistic estimates of detection probabilities.

On the other hand, in reality a fraction of the contamination will precipitate in the mains. This may either reduce the detection probability as *E. coli* is removed from the water phase or increase the detection probability due to longer retention and periodic resuspension.

Contamination events in the distribution network are the cause of a large proportion of outbreaks of illness through community water supply systems; in the 44 community supply outbreaks that occurred in the US between 1991 and 1998, 17 (39%) of the outbreaks were attributed to a contamination event in the distribution system (Craun et al., 2003). The low probability of detection of faecal contamination is in line with the outbreak reports that indicate that the outbreak is detected through consumer taste complaints or illness reports rather than water quality monitoring (Fernandes et al., 2006; Proctor et al., 1998; Hrudey and Hrudey, 2004). When consumer illness is reported and water quality monitoring is directed to the outbreak location, traces of a contamination event can sometimes be found from *E. coli* monitoring (Huisman and Nobel, 1981) or from pathogen monitoring (MacKenzie et al., 1994) or retrospective analysis of treatment performance data (e.g. turbidity, MacKenzie et al., 1994) or events that may have led to contamination (O'Connor, 2002).

Further investigations of optimisation possibilities will include selection of representative sampling sites by:

- stratification based on e.g. mains materials, age of pipes and other variables that may influence risks of contaminations (Speight et al., 2004; Narasimhan and Brereton, 2004);
- covering a maximum fraction of the flow pattern, using algorithms developed by Lee and Deininger (1992) and optimised by Harmant et al. (1999);
- probability of detection of a contamination event (e.g. reservoirs).

Another option that will be evaluated is to increase the frequency of sampling, given the transient nature of contamination events in distribution networks. (Semi-)continuous

on-line detection methods (when they become available) could provide the highest probability of detection.

In the current grab-sample monitoring programmes, it is better to increase the frequency of sampling than to increase the volume of sampling and analysis, assuming a non-random distribution of contamination events in distribution systems (Haas, 1993). However, the volume of each sample should be as large as possible without significantly increasing the costs of collecting and analysing these larger samples. The optimum can be calculated simulating different sampling programmes with equal costs.

Optimisation of monitoring programmes to increase the probability of detecting faecal contaminations of drinking water distribution systems should not lead to a disproportionate increase of operational costs. Monitoring of drinking water serves as a verification of the effect of systems preventing these contaminations and the majority of funds should therefore be directed to these preventive systems.

5. Conclusions

- Although the case study has only evaluated 12 out of many possible simulated contamination events, the results make a reasonable case for a low probability of detecting even severe faecal contaminations of drinking water mains with standard monitoring programmes.
- The mean probability of detecting a 16 h contamination of mains with untreated domestic sewage ($1\,l\,h^{-1}$) containing 1×10^8 *E. coli* in the model system was 5% (SD 5%, median 6%, range 0–14%, $n = 9$).
- Identical contaminations of the finished water, a trunk main or a reservoir are more likely to be detected (mean probabilities 55–65%, SDs 43–45%).
- In reality, these detection probabilities will be even lower as the case study assumed no inactivation or clustering of *E. coli*, 100% recovery efficiency of the *E. coli* detection methods and immediate mixing of contaminations in mains and reservoirs.
- The study also showed that simple adjustments such as a more homogeneous spreading of sampling moments can increase the probability of detection.
- More complex improvements may be possible as well, such as determining the best time and place for sampling; this will be evaluated in a follow-up study.

Acknowledgements

This research was conducted within the framework of the Joint Research Program (BTO) of The Netherlands water supply companies. The authors are especially indebted to water company Vitens for providing the data for the hydraulic model and the sampling programme.

REFERENCES

Ashbolt, N.J., Grabow, W.O.K., Snozzi, M., 2001. Indicators of microbial water quality. In: Fewtrell, L., Bartram, J. (Eds.), Water Quality: Guidelines, Standards and Health. IWA Publishing, London, UK, pp. 289–316.

Bjergbæk, L.A., Roslev, P., 2005. Formation of nonculturable *Escherichia coli* in drinking water. J. Appl. Microbiol. 99, 1090–1098.

Craun, G.F., Calderon, R.L., Nwachuku, N., 2003. Causes of waterborne outbreaks reported in the United States, 1991–1998. In: Hunter, P.R., Waite, M., Ronchi, E. (Eds.), Drinking Water and Infectious Disease. Establishing the Links. CRC Press/IWA Publishing, London, UK.

EPA, 2000. EPANET 2.0 (<www.epa.gov/ORD/NRMRL/wswrd/epanet.html>; last accessed November 6, 2006). Users manual written by Rossman, L.A. (2000). Report EPA/600/R-00/057, US Environmental Protection Agency, Cincinnati, OH.

European Council, 1998. Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption.

Fernandes, T.M., Schout, C., De Roda Husman, A.M., Eilander, A., Vennema, H., Van Duynhoven, Y.T., 2006. Gastroenteritis associated with accidental contamination of drinking water with partially treated water. Epidemiol. Infect. epub 17 November, pp. 1–9.

Haas, C.N., 1993. Microbial sampling: is it better to sample many times or use large samples? Water Sci. Technol. 27 (3–4), 19–25.

Hambsch, B., 1999. Distributing groundwater without a disinfectant residual. J. Am. Water Works Assoc. 91, 81–85.

Harmant, P., Nace, A., Klene, L., Fotoohi, F., 1999. Self inspectorate of potable water distribution systems (Auto-contrôle optimisé dans les réseaux de distribution d'eau potable). L'eau, l'industrie, les nuisances 225, 49–52.

Huisman, J., Nobel, P.J., 1981. Some epidemiological data on the effects of the faecal contamination of drinking-water in the Scheepvaartkwartier residential area in Rotterdam in March 1981 (Enkele epidemiologische gegevens over de gevolgen van de faecale drinkwaterverontreiniging in het Scheepvaartkwartier te Rotterdam in maart 1981). H_2O 14 (26), 642–646.

Hrudey, S.E., Hrudey, E.J., 2004. Safe Drinking-Water. Lessons from Recent Outbreaks in Affluent Nations. IWA Publishing, London, UK.

Lee, B.H., Deininger, R.A., 1992. Optimal locations of monitoring stations in water distribution system. J. Environ. Eng. 118 (1), 4–16.

MacKenzie, W.R., Hoxie, N.J., et al., 1994. A massive outbreak in Milwaukee of Cryptosporidium infection transmitted through the public water supply. N. Engl. J. Med. 331, 161–167.

Medema, G.J., Schijven, J.F., 2001. Modelling the sewage discharge and dispersion of *Cryptosporidium* and *Giardia* in surface water. Water Res. 35 (18), 4307–4316.

Medema, G.J., Ketelaars, H.A.M., en Hoogenboezem, W. (Eds.), 2001. *Cryptosporidium* and *Giardia*: Presence in Sewage, Manure and Surface Water with Swimming and Drinking Water Application (*Cryptosporidium* en *Giardia*: Voorkomen in Rioolwater, Mest en Oppervlaktewater Met zwem- en Drinkwaterfunctie). RIWA, Amsterdam, 171pp.

Narasimhan, R., Brereton, J., 2004. Developing a Bacterial Sampling Plan. Guidance Manual. AWWA Research Foundation, Denver, USA, 68pp.

O'Connor, D., 2002. Report of the Walkerton Inquiry: The Events of May 2000 and Related Issues. Part One: A Summary. Ontario Ministry of the Attorney General, Toronto.

Payment, P., 1999. Poor efficacy of residual chlorine disinfectant in drinking water to inactivate waterborne pathogens in distribution systems. Can. J. Microbiol. 45, 709–715.

Proctor, M.E., Blair, K.A., Davis, J.P., 1998. Surveillance data for waterborne illness detection: an assessment following a massive outbreak of *Cryptosporidium* infection. Epidemiol. Infect. 120, 43–54.

Rossman, L.A., Clark, R.M., Grayman, W.M., 1994. Modeling chlorine residuals in drinking water distribution systems. J. Environ. Eng. 120 (4), 803–820.

Speight, V.L., Kalsbeek, W.D., DiGiano, F.A., 2004. Randomized stratified sampling methodology for water quality in distribution systems. J. Water Resourc. Plann. Manage. 130 (7/8), 330–338.

Van der Kooij, D., Van Lieverloo, J.H.M., Schellart, J.A., Hiemstra, P., 1999. Maintaining quality without a disinfectant residual. J. Am. Water Works Assoc. 91 (1), 55–64.

Van Lieverloo, J.H.M., Medema, G.J., Van der Kooij, D., 2006. Risk assessment and risk management of faecal contaminations in drinking water distributed without a disinfectant residual. J. Water SRT–Aqua 55, 25–31.

Vasconcelos, J.J., Rossman, L.A., Grayman, W.M., Boulos, P.F., Clark, R.M., 1997. Kinetics of chlorine decay. J. Am. Water Works Assoc. 89 (7), 54–65.